Special Report 84-2

SOME CRITICAL FACTORS THAT LIMIT THE EFFECTIVENESS OF MACHINE INTELLIGENCE TECHNOLOGY IN MILITARY SYSTEMS APPLICATIONS

S. D. Harris* and J. M. Owens

September 1984

NAVAL AEROSPACE MEDICAL RESEARCH LABORATORY

PENSACOLA FLORIDA

Approved for public release; distribution unlimited.

Some Critical Factors that Limit the Effectiveness of Machine Intelligence Technology in Military Systems Applications

S. D. Harris* and J. M. Owens

Naval Air Systems Command
61153N WR04210.WR0421001.6142

Reviewed by

F. E. Guedry, Ph.D.
Senior Scientist

Approved and Released by

Captain W. M. Houk, MC USN
Commanding Officer

September 1984

*Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061

Naval Aerospace Medical Research Laboratory
Naval Air Station, Pensacola, Florida 32508

SUMMARY PAGE

THE PROBLEM

Recent advances in machine intelligence (MI) technology have been widely touted as a panacea for many difficult military systems problems. However, a number of factors that limit the usefulness of the technology in military applications become apparent upon perusal of the available data that reflect the performance of operational MI systems. This report serves to delineate some important problems and issues that require the attention of the research community if machine intelligence technology is to become a viable component in defense systems.

FINDINGS

Machine intelligence technology is not the product of a recent breakthrough in computer system design as many have been led to believe. The fundamental concepts, and many of the underlying problems in the design of today's intelligent machines, were recognized fifty years ago, and earlier. Research efforts in diverse areas, some spanning decades, have culminated in the current state of the technology. However, numerous and often perplexing methodological problems currently challenge system development and thwart the widespread and successful application of the technology in critical areas.

The issue discussed in this paper arise from the failure of MI technology to confront and effectively solve many of the complex problems inherent in military systems applications. Most critical is the fact that performance criteria for intelligent machines are vague and inadequate, or completely unspecified in some instances. It is often impossible to ascertain from the reported data exactly what a system is supposed to do and how well its functions are performed. The problem of performance assessment is one of several difficult issues discussed in this report.

RECOMMENDATIONS

There are no ready solutions to the problems outlined in this report. Systematic research and development that concentrates on well-defined military systems applications and that simultaneously attends to the issues raised here would represent an important and necessary step in the evolution of MI technology.

Acknowledgments

The authors wish to express thanks to Yee-Yeen Chu, Perceptronics, Inc., Larry Egan, Logicon, Inc., and Mike Fehling, Advanced Information and Decision Systems, Inc. for their responsive support of the investigators on a very short notice. A special debt of gratitude is owed Mrs. Darlene Schaeff and her staff, Naval Reserve Readiness Command Region Six, for their

perseverence and professionalism in attending to a number of complex logistical problems. Their dedication to get the job done made it possible for the authors to complete the data and information gathering required for this report. Thank you, Mrs. Schaeff.

TABLE OF CONTENTS

INTRODUCTION

By many accounts, the most significant advance in information processing technology during recent years has been the emergence of machine intelligence (MI) as an engineering tool. This event was precipitated by an almost classical confrontation between scientific research conducted in academic laboratories, and the "real world" needs for system applications to augment human performance. The purpose of this report is to provide a brief critical assessment of machine intelligence technology, with particular attention to the special problems that arise in military systems applications. A literature review in several related but distinct domains, and discussions with government research and development managers and representatives from industry and academia served to identify some of the knowledge gaps that impede progress in developing MI for military medical and aviation systems applications. This report summarizes the most salient and critical of these problems and issues.

Historical Perspective

The history of research in MI can be traced to the early 1930's when Godel (7) formally delineated the inherent difficulties of computation with axiomatic systems. In fact, discussions of some of the issues involved in producing intelligent machines can be traced even further back in time (11). For instance, one vital research area that currently enjoys considerable attention concerns the use of symbolic logic invented by Aristotle (10). In 1943, Post (18) formulated the seminal concept that all forms of computation were formally equivalent, and posited a theory of computation called the production system model. In the middle 1950's, a number of researchers interested in psychological processes began attempts to model thinking as a computational process (17). In the best tradition of academe, they chose problem domains that were difficult enough to challenge both the theoretical positions that were espoused, and the methods at their disposal. Their methods involved the use of computers to implement and test their computational theories of thinking, and therein lay a revolutionary idea.

If a computer program could be written that adequately tested a theoretical depiction of human problem solving, and if the theory were correct, then the working program would be potentially useful as a real problem solving tool in either augmenting human performance or perhaps in displacing humans in some realms. In the ensuing decade and a half, research activity waxed and waned in a number of areas including, visual and speech perception, planning, language understanding, and, to some extent, machine learning. In addition to initial formulations and developments, certain tools began to evolve to aid researchers in their tasks of constructing ever larger and more complex theories. Notably, a programming language called LISP was invented by John McCarthy (13), and rapidly became the standard for MI research. The technology also benefited

substantially by concomitant major improvements in computing hardware--faster computers with more memory--that made it possible to explore rapidly the limits of the complex theories.

Recent Developments

Many theoretical, methodological and technical developments occurred during the early 1970's to advance the state of MI technology. Developments in diverse areas, such as graph theory, search algorithms, new programming languages and computer support environments, machines designed to maximize the performance of MI systems, psychological theories of learning, memory organization, perception and problem-solving, to name a few, were important in furthering the technology. One of the more significant events was the publication by Newell and Simon (17) of a formal model of human problem solving based on Post's tenets.

Partly in response to the catalytic effect of Newell and Simon's work, several ideas matured to the point that some scientists turned their attention from the simple laboratory tasks of traditional MI research to substantially more difficult real-world problems. The turn toward the design and engineering of systems with practical implications is best exemplified by the development of systems designed for medical applications, and, of particular concern for present purposes, for military tactical situation assessment. These systems, and their progeny have come to be called "expert" systems or "knowledge-based" systems. However complex and abstruse the architecture and data structures of these systems, they are all variations of the Post production system concept.

A second development of consequence during the early 1970's, was the invention of PROLOG, a new computer programming language especially well-suited for expressing and proving logical propositions (3). PROLOG is important because it is the first computer programming language to embody a formal mathematical theory of computation. Hence, it provides a tool that is much more efficient and effective than other programming languages for modeling thinking as a computational process.

Advances in the area of computer hardware design were also critical to the evolution of MI technology. The first four "generations" of computers, beginning with Eniac and including all commercially available computers today, were based upon an idea introduced by John von Neumann in 1946 (16). The generations were characterized primarily by increasing reliance upon improvements in solid-state electronic technology that provided increasing miniaturization of circuitry. Faster processors and more memory were made available in smaller volume packages. But, the fundamental systems architecture remained essentially unchanged. The current, highly publicized revolution in computer technology stimulated by the Japanese fifth generation computer initiative (6) and the resulting programs in the U.S., such as the Defense Advanced Research Projects Agency (DARPA) strategic computing program, are the products of a radically different

concept known as "non-von Neumann" architecture. New machines are being designed to optimize the execution of logic programs rather than high-speed arithmetic. Logic programs are the pragmatic expression of the Post production system formalism.

Diverging Design Approaches

Toward the end of the last decade, the research fields concerned with the development of intelligent machines split into two camps: artificial intelligence (AI) research on the one hand, and cognitive science (CS) on the other. The AI and CS disciplines are represented in the U.S. by the American Association for Artificial Intelligence (AAAI), and the Cognitive Science Society, respectively. The differences between the two disciplines are both philosophical and methodological.

Cognitive Science researchers tend to concern themselves with discovering principles of human intelligence, often through controlled experiments with humans. But, as Michalski, Carbonnel, and Mitchell (14) noted, computers are to human intelligence as airplanes are to birds--there are fundamental similarities, but also a number of fundamental differences. In that vein AI researchers declare that machines should not be constrained to function as humans, rather, machine intelligence research should lead to new forms of intelligence optimized to exploit the strengths of machines.

The construct that unifies the two disparate fields is the computer as a research tool. The criterion for success in AI is a working program. The criterion for success in CS is a program that works in a manner that has a plausible basis in the results of experimentation with humans; in other words, the program must simulate human intelligence. The differences in the goals and approaches of AI vs. CS proponents, therefore, can often result in radically different types of systems. Although the goals of AI and CS have differed in the past, a convergence of efforts in key areas would undoubtedly benefit the production of more viable operational systems.

By far the major impetus for applied machine intelligence has come from the AI field. However, there is reason to believe that the approach proffered by AI will enjoy limited success unless and until there are advances in underlying theories of human intelligence. The most telling evidence is in the reported problems regarding the design of human-machine interfaces. For the very large class of applications that necessarily entail intimate interaction between human and machine intelligence, the communication problem is serious. It seems that no matter how effective an AI system is, under some conditions both the results of its processing and some explication of its rationale must be presented to a human in a manner that makes sense to the human; i.e., the system must appear to think about the problem in a manner that is human-like. Future attempts to develop practical

systems with which humans must interact will necessarily require the methods of both AI and CS and, consequently, improved collaboration between the disciplines.

Commercial Prospects

Evidence of the vitality of the AI industry can be found in several places, including recruiting materials published by several large corporations and government agencies seeking computer science professionals with experience in AI. Membership in the Cognitive Science Society and the American Association for Artificial Intelligence is also rapidly increasing. Further, market surveys have been published that project an industry on the order of $250 billion by 1990 (Defense Electronics, August, 1983).

The focusing of increasingly precious human and computational scientific resources on the development of operational systems has had an unfortunate side effect, i.e., the depletion of resources available for purely theoretical research. According to one source (4), there are only about 60 people in the world with high-level expertise in the development of knowledge-based systems. However, the same sources that tout that the shortage has reached crisis proportions also claim to be among the few who have the special talent. In any event, there is an undisputed shortage of scientific research currently being performed. The same economic forces that attract all engineering and technical people away from scientific endeavor have apparently begun to impact AI as well.

AN OVERVIEW OF PRODUCTION SYSTEMS

As previously discussed, the foundation of many, if not most, of the programs that comprise operational AI systems is the Post production system concept. The remainder of this section of the report introduces the idea of a production system and some elements of the programming language PROLOG to illustrate the most elementary type of "reasoning" inherent in many of the working systems. The following section then turns to a discussion of several of the critical issues that deserve attention if AI is to realize its potential in military systems applications. Overall the intent is to highlight the following questions: What are intelligent systems? How are they constructed? What are the unknown, limiting factors in their development?

Production Rules and Facts are the Elements of Knowledge

A production rule is a data structure comprising two parts: a condition part and a conclusion part. A rule encodes a logical implication. A fact is just a degenerate rule, having a conclusion part that encodes some kernel of truth without any qualifications.

The general form of a production rule can be represented as

((conclusion) <- (condition)),

read "condition implies conclusion." The general form of a fact can be represented as

((fact)).

A complete production system consists of two components: a "knowledge base" made up of many (hundreds or thousands) such rules and facts, and a control mechanism that searches the knowledge base for information that will solve a problem.

In a mathematical sense, a problem may be specified to the system as a theorem to be proved. In a more general sense, a problem specification may be a goal statement that symbolizes a desired state in the knowledge base. The control strategy operates in an iterative fashion, selecting a production rule, applying that rule, then selecting a new rule and iterating. The selection is made by comparing the conditions specified in a rule with the facts in the knowledge base. If the condition(s) specified in the rule are consonant with the fact(s) in the knowledge base, then the rule is applicable. The system then selects from among all applicable rules the one rule to apply next. In a pure production system, the truth of the conclusion of an applied rule is simply taken into account by the control strategy in its search for information relevant to the computational problem at hand. Because data (facts) and programs (rules) are encoded in the knowledge base in the same fundamental form, it is possible for a program to modify itself. That is, the conclusion of a rule may be that a new fact or rule should be added to the knowledge base. Such outcomes are called side effects. When side effects result in modification of the control mechanism's own logic, the system is said to "learn."

Heuristics Guide the Search for Truth

AI control mechanisms usually entail a complex combination of elementary, systematic and mathematically provable search strategies, and less well-founded, domain-specific search techniques called heuristics. In the simplest form, a heuristic might be a rule that causes the control mechanism to search a restricted part of the knowledge base first for applicable rules or relevant facts before trying other sections of the knowledge base. If the heuristic is effective, then the overall performance of the system should improve.

In most operational AI systems, there is a certain ambiguity included in the structure of the rules. Thus, the <- symbol might be read "suggests" rather than "implies," and a quantity (similar to a subjective probability estimate) might be assigned to the rule to represent the strength of the suggestion. A fairly complex calculus of evidence must then be introduced into the heuristic search mechanism to exploit the encoded

"knowledge" about uncertainty (5). The process of designing the knowledge base and the search mechanism, which usually entails extracting facts, implications, suggestions, and other heuristics from the brains of experts in a particular domain, and encoding that information in some machine representation, constitutes the major design activity of the so-called "knowledge engineers." These individuals specialize in designing MI systems for real-world applications.

<u>An Example: Paris' Amorous Inclinations</u>

Consider the following example of an elementary system that knows about "love."

```
((female Hera))
((blonde Hera))

((female Helen))
((brunette Helen))

((male Paris))

((lovable *someone) <- (female *someone)
                       (blonde *someone))

((loves Paris *someone) <-  (lovable *someone))
```

A fact, such as ((female Hera)) can be interpreted as "Hera is a female." A rule, such as

```
((lovable *someone) <- (female *someone)
                       (blonde *someone))
```

can be interpreted as "If someone is both female and blonde, then that person is lovable." Alternatively, the rule may be interpreted as the fact that someone who is both female and blonde <u>implies</u> (<-) that that person is lovable. The reader is invited to ascertain from the facts and rules represented in the example the answer to the question "Whom does Paris love?" A slightly less pedantic example should suffice to convey the power of the production system concept, and bring us to the point of immediate interest, i.e., military systems applications.

<u>An Example: The F-14 Tactical Decision Aid</u>

The F-14 is a U.S. Navy fighter-type aircraft designed for all-weather fleet air defense. It operates from an aircraft carrier, and is in the Mach 2.3+ class. It carries a variety of weapons and is specialized for long-range air-to-air combat. One mission of the F-14 is Combat Air Patrol (CAP). During a typical mission, the F-14 is launched from a carrier and proceeds to an assigned position 150 miles or so from the task force. While circling the patrol station, the crew will receive secure voice communications from an Airborne Early Warning (AEW) aircraft (an E-2C) to investigate an unidentified air target flying in the

direction of the task force. The fighter will receive steering information until it acquires the target or targets (called a "threat cloud") on its own radar and assumes control of the intercept.

The following,

```
((offensive) <- (!(posture (pla offensive)))
                 (!(c1))
                 (!(c2))
                 ((c3))
                 ((c4))
             (replace posture (pla offensive)))
```

is a rule excised from a tactical decision aid (TDA) under development for the F-14 aircraft (12). The rule may be interpreted as follows:

> If the current posture is not (!) offensive, and if parameters (C1) and (C2) are false while parameters (C3) and (C4) are true, then assume an offensive posture.

Parameter (C1) is true if the incoming threat has penetrated an imaginary envelope around the carrier task force being defended, else it is false; (C2) is true if fuel on-board the fighter is "critical," else false; (C3) is true if the number of potential targets in the incoming threat cloud is less tnan the number of available Phoenix missiles aboard the fighter, else false; (C4) is true if the cumulative threat cloud lethality index (a measure of the danger represented by the make-up of the threat) exceeds a certain threshold, else false. To reinterpret the rule,

> The system will recommend an aggressive tactical stance, with related maneuvers and guidance for the intercept, if the incoming threat has not yet penetrated the envelope of vulnerability, fuel aboard the fighter is uncritical, the number of on-board missiles exceeds the number of targets, and the target is considered highly lethal.

The F-14 TDA is designed to process data from the tactical information display (TID) in the cockpit, and the radar, to assess the tactical situation according to rules such as the one illustrated above, and to recommend maneuvers to the crew to achieve the best tactical advantage during air combat maneuvering. The system aggregates the information from many sources, including the air data computer and fuel gauges, in addition to the TID and radar. This type of information processing is referred to as "information fusion."

The rule presented above is invoked by the rule selection mechanism during an intermediate step in the process. The output, which recommends an "offensive" posture, has a

significant impact on subsequent computations of recommended turns, selection of target subsets to engage, and other processes.

The examples of Paris' amorous inclinations and the F-14 TDA, have served to introduce the elements and functioning of a production system. Rather than continue with a more formal and necessarily lengthy treatment of search strategies and the calculus of uncertainty, an example of a dialogue between a sophisticated user and an early "successful" expert system known as MYCIN (20) will be presented at this point. From this example, the reader may individually judge whether the expert system concept, based as it is on a disarmingly simple idea, does indeed produce useful results and evidence of "intelligent" behavior.

An Example Consultation with MYCIN

MYCIN was an early AI system designed to aid a physician in diagnosing certain types of infectious diseases and prescribing antimicrobial therapy. The inset below illustrates a portion of a consultation between a physician and MYCIN concerning a fictitious patient, Jeremiah Sample. In the dialogue, the physician's entries follow the double asterisk (**) prompt symbol. The system's entries are capitalized. The entry 'Rule' is a query for the system to state the rule that it is currently attempting to employ. The entry 'Why' is a request for the system to provide an explanation of how it arrived at its current line of reasoning.

22) WAS ANY CHEMOTHERAPY RECENTLY BEGUN SPECIFICALLY TO COMBAT ORGANISM-1? ** No 23) IS JEREMIAH SAMPLE (PATIENT-1) A BURN PATIENT? ** Rule
RULE047

IF:

1) THE SITE OF THE CULTURE IS BLOOD, AND

2) THE IDENTITY OF THE ORGANISM IS NOT KNOWN WITH CERTAINTY, AND

3) THE STAIN OF THE ORGANISM IS GRAMNEG, AND

4) THE MORPHOLOGY OF THE ORGANISM IS ROD, AND

5) THE PATIENT HAS BEEN SERIOUSLY BURNED

THEN:

THERE IS WEAKLY SUGGESTIVE EVIDENCE (.4) THAT THE IDENTITY OF THE ORGANISM IS PSEUDOMONAS

** No 24) IS THE PATIENT'S ILLNESS WITH ORGANISM-1 A HOSPITAL-ACQUIRED INFECTION? ** Yes 25) DO YOU CONSIDER JEREMIAH SAMPLE (PATIENT-1) TO BE A COMPROMISED HOST? (ALCOHOLIC, IMMUNOSUPPRESSED, ETC.): ** No 26) HAS JEREMIAH SAMPLE (PATIENT-1) HAD ANY GU SURGERY OR MANIPULATION OF THE URINARY TRACT? (E.G., INDWELLING CATHETER, CYSTOSCOPY, ETC.) ** Why

I AM TRYING TO FIND OUT ABOUT THE IDENTITY OF THE ORGANISM. I ALREADY KNOW THAT: THE PORTAL OF ENTRY OF THE ORGANISM IS UPPER-GU THE MORPHOLOGY OF THE ORGANISM IS ROD THE STAIN OF THE ORGANISM IS GRAMNEG THE SITE OF THE CULTURE IS BLOOD THEREFORE, IF: 1) THE PATIENT HAS NOT HAD A GENITO-URINARY MANIPULATIVE PROCEDURE, AND 2) URINARY-TRACT-INFECTION IS NOT A PROBLEM FOR WHICH THE PATIENT HAS BEEN TREATED THEN: THERE IS SUGGESTIVE EVIDENCE (.6) THAT THE IDENTIFY OF THE ORGANISM IS E.COLI (RULE156) ETC.

The transcription is intended to convey the general level of interaction between user and system. In looking over the example, the reader should note several things that are characteristic of current AI systems. The dialogue is very specific to the problem domain, i.e., infectious diseases. At least to a naive observer, the system appears to "know" a lot about infectious diseases. There is some provision in the knowledge base for representing and using estimates of "certainty" in the form of suggestive evidence on a scale of Ø to 1. There is a very subtle personification evident in the dialog. At one point MYCIN says "I am trying . . .". Finally, the dialogue itself is somewhat stilted and stylized. The natural language capability of MYCIN was rather limited.

THE PROBLEMS AND ISSUES

As mentioned at the outset, one major objective of applied MI technology is to augment human performance. Human performance can be thought of as comprising two types: procedural, e. g., executing sequences of steps correctly to achieve some desired objective; and cognitive, e. g., selecting the appropriate objective to pursue. It is probably accurate to say that traditional human factors engineering has as a goal the enhancement of procedural performance through improved design of the superficial characteristics of the human-system interface. Properly designed keyboards and furniture, for example, improve productivity (8).

On the other hand, defining and improving cognitive performance, such as enhancing memory, or the quality of decisions has proven a much more elusive goal. It is in this area, improved cognitive performance, that MI has shown great promise. The primary objective of the knowledge-based systems

has been to aid the problem-solving performance of very sophisticated users in limited domains. Perusal of the many recent articles that discuss AI applications will almost inevitabily turn up references to MYCIN, as well as some allusions to systems that have discovered multi-million dollar mineral lodes, or systems that discover new mathematical concepts. In point of fact, the list of operational systems is surprisingly short, and their achievements are usually quite circumscribed. To date, the promise of MI has been thwarted by several unresolved problems. Some of the unresolved problems and issues that have emerged from the several attempts to build operational systems are discussed next.

<u>Artificial Intelligence is Undefined</u>

One of the more disconcerting aspects of the task of evaluating applied AI technology is a certain nebulosity underlying the extensive technical literature. Consider the problem of defining "artificial intelligence." The term "artificial" obviously refers to "human-made" but the term "intelligence" remains essentially undefined. Most text books simply assert that AI is the discipline concerned with developing intelligent machines, and then turn to discussions of techniques and concepts related to the organization of AI systems. Typical of the genre is the definition proffered by Rich (19):

> <u>Artificial Intelligence</u> (A.I.) is the study of how o make computers do things at which, at the moment, people are better....So, for at least the next few years, this definition should provide a good outline of what constitutes artificial intelligence, and it avoids the philosophical issues that dominate attempts to define the meaning of either artificial or <u>intelligence.</u>

We disagree with Rich. One of the most serious threats to the integrity of AI as a scientific discipline, in our opinion, is the unwillingness of its oponents to keep the question "What is intelligence" in the forefront of the research issues. Without benefit of a formal definition, operational definitions abound! An operational definition that most often seems to apply goes something like this:

> If the machine's designer declares that it is intelligent, and if the human user cannot predict the behavior of the system within the constraints of an operational milieu, and if the user is impressed with the performance of the system, then it is intelligent, for all practical purposes.

This operational definition would probably scratch an open sore if brought to the attention of most serious AI researchers. It seems that if one considers oneself an AI scientist, the question "What is artificial intelligence?" is viewed as evidence of disbelief that AI is "for real," as declared on a bumper

sticker recently published by the American Association for Artificial Intelligence (AAAI). From the standpoint of a potential consumer of AI, the problem is to discriminate between the real thing, and imposters. Without criteria to make that distinction, the AI field, and the consumer, invite some persons to exploit the hype (e.g., bumper stickers and luggage tags) being generated about the field in popular and quasi-technical media. Anecdotal stories about the problem are rampant!

There is an important distinction between the theory and the practice of AI. It is clear, after some rumination, that AI is firmly grounded on well-established mathematical and engineering principles. Further, many of the less well-founded ideas are subject to intense scrutiny of rigorous, skeptical scientists. Those ideas that survive will undoubtedly be innovative solutions to long-standing problems in computer science. However, it is often a very long intuitive leap from scientific principle to real-world application. Unfortunately, engineering rigor often seems to be subtly supplanted by more *ad hoc*, serendipitous solutions to problems. The danger with such an approach is that systems that appear to work under benign conditions may fail catastrophically under conditions in the operational milieu (9).

New Engineering Tools are Needed

The AI literature is replete with references to a concept known as "knowledge engineering" (KE). And, more recently in the literature of human factors, one sees the term "cognitive engineering." Neither of these terms is well-defined, and we are skeptical that they represent a true engineering discipline, except in the loosest sense of the term. However, the terms do convey an essential idea: i.e., the limiting factor in the design of many operational AI systems has to do with capturing the knowledge and cognitive skills of human problem solvers in peculiar application domains, such as medical diagnosis or air combat, and encoding that knowledge in a computer program. The validity, reliability, specificity, and internal consistency of the opinions of the expert problem solvers must be assessed before one can be reasonably assured of success in an attempt to build a system.

An evaluation of expert opinions is a complicated matter as there are no widely accepted methods for extracting and assessing knowledge from human experts. The methods that are available for systematically performing KE are often proprietary to producers of the systems. It can be surmised that the methods must be very effective and provide their owners with competitive advantages. The alternative conclusion, that the methods are either nonexistent or untried and deserve concealment to protect a fundamental weakness in the KE process is not very satisfying. It appears doubtful that any of the commercially available methods has, as yet, been used to build a successful application system. A recent report by Waterman and Hayes-Roth (21) addresses some of these issues in detail.

Knowledge Representation is Often Atheoretical

Most of the available AI systems have been experimental in nature, developed to test scientific hypotheses and concepts. The representation of knowledge has been influenced significantly by the peculiar encoding scheme preferred by the knowledge engineer. Very little research has been performed that bears on the question of the comparative merits of the many alternative representations. The choice of one scheme over another is often a matter of speculation rather than the result of rigorous and systematic analysis.

The representation chosen for a given problem, then, may be more a function of a best guess on the part of the system's programmer than a result of systematic analysis. From the standpoint of research in AI, the choice of a representation schema may be based on a scientist's "hunch," intended to test a theory. In the realm of scientific inquiry, theories and hunches lead to significant progress, but the systems developer should not be misled; there are numerous theories, but there is no consensus.

Performance Criteria are Vague

Performance data on operational AI systems are nearly nonexistent. In fact, many professionals who purport to develop "high performance" systems seem genuinely puzzled by the query "Can I see the performance data?" Often there are no data. The "success" of a system is in the mind of its builder. This situation is obviously unsatisfactory from the standpoint of military applications although it may not be especially critical for less demanding systems.

The fundamental question faced by the system designer is whether a system can consistently provide adequate solutions to the problems at hand. The solutions proffered by an AI system may be reasonable to one human, and unsatisfactory to another, more sophisticated and better informed human. The sufficiency of the system is inextricably bound to the characteristics of the user.

It is unfortunate that the assessment of the performance of AI systems has been primarily from the perspective of AI research needs related to architectural considerations, theoretical models, improved search proficiency, etc. Tests of AI systems in operational settings are almost nonexistent. Those reports that document the results of operational tests have repeatedly identified the human-machine interface as a limiting factor in the effectiveness of an AI system. The problem goes beyond the traditional issues in human-machine interface design. In addition to the problems related to visual display formatting, query language design, etc., the AI system introduces new problems, such as the explication of the rationale employed by the system in reaching its conclusions and recommendations (1).

Uncertainty is a Critical Dimension

If one is contemplating whether to develop an AI-based system, it appears reasonable to ask whether the problem is well-specified, or is rife with uncertainties. The MYCIN system, discussed earlier, introduced one approach to coping with uncertainty in the representation code. The problem domain to which MYCIN was addressed is characterized as ill-conditioned because the issues are so complex that even world-class experts may disagree on the diagnosis. On the other hand, consider a truly "operational" system, MACSYMA.

MACSYMA is a system comprising a variety of mathematical symbolic manipulation and integration capabilities. MACSYMA's forte is the integration of mathematical functions that entails not only computing the integral, but also figuring out the symbolic integral of a given algebraic expression. MACSYMA can be described as extremely successful (1). One of the many reasons for its success may well be the inherent deterministic nature of the problems to which it is addressed. The integration of mathematical functions is a difficult problem but it is undoubtedly more straightforward than medical diagnosis. The lesson is simple. If solutions and information sources can be identified in the problem domain with certainty, then the knowledge-based approach can probably be exploited readily. If uncertainty prevails, either in the operational information sources, or in the decision-making/problem-solving strategies that experienced humans employ in the operational milieu, then an attempt to automate a problem-solving system based on AI will necessarily encounter the problem of representing uncertainty, supplanting logical implication with rational suggestion in the fundamental coding of knowledge.

Workload Assessment Techniques are Needed

If the intent of a system is to enhance the "cognitive performance" of humans, it would seem reasonable to start the design of the system with an appeal to certain considerations from the human factors discipline. The first thing a human factors specialist is wont to do is perform a "task analysis." Although a task analysis will generate detailed descriptions of the behaviors expected of the humans in the system, the information sources, the decisions, the control actions, etc., these descriptions are often qualitative. There are no generally accepted metrics of information or decision-making that can quantify cognitive performance in an operational milieu. In essence, what is needed is a methodology for measuring the "cognitive workload" currently imposed on the humans for which the AI system is intended as a performance augmentor. Such a measure would be useful for diagnosing an existing workload problem and for assessing the effectiveness of a proposed solution.

The production system model of human problem solving (17) suggests one approach to assessing cognitive workload. One product of the knowledge engineering process is an encoded knowledge base, a collection of facts and rules extracted from human experts. Although it is unclear whether the experts actually solve the problems by invoking the rules in the operational setting (17), it may be reasonable to consider the knowledge base as a model of the information processing required of the humans. Research on human performance (15) suggests that the maximum number of rules that might be applicable at any given moment must be relatively small, certainly less than ten. Further, the complexity of the rules might be similarly constrained to fewer than ten conditions that must be simultaneously tested for a conclusion to be suggested or implied. Thus, cognitive workload might be assessed as the average number of rules invoked per unit time or per critical decision, or as the average number of conditions per rule.

Stress Affects Human-Machine Dynamics

Military tasks are stressful. The impact of this stress on short-term memory capacity, decision quality, and other aspects of cognitive performance is complex and difficult to assess. The production system model of Newell and Simon (17) represents human cognition as a limited capacity system that iteratively produces one of a set of states; the peculiar state is chosen in response to evidence derived from the senses and from memory. Evidence from some sources may be given more weight than that from others, and some states may be positively biased, independently from available evidence. The system's states correspond to states of the world (the task environment), control or bias the emission of behaviors, and influence accession to memory. The parameters that govern the various capacities and biases are changed by changes in the general state of the nervous system; these changes may be produced by stress. As a result, capacities are likely to be reduced and biases distorted under stressful performance conditions (2).

The implications of stress effects on human performance in operational military systems and, consequently, on the efficacy of MI technology developed for such systems are especially problematic at this time. Human behavior and human-machine relations are subject to change under stressful conditions and these factors must be taken into account in the MI design process. Furthermore, systems design criteria for MI technology must be empirically established to include the system's ability to monitor and compensate for system or human behaviors out of reasonable bounds and to execute appropriate alerting or corrective actions. The challenges for MI technology development for operational military applications and the associated research issues are clearly numerous.

SUMMARY AND RECOMMENDATIONS

A considerable proportion (perhaps 25 percent) of the progress to-date in building operational AI systems has revolved around the development of diagnostic aids for physicians. A cursory example of the first "successful"medical diagnosis system was presented to illustrate some aspects of the design and behavior of expert systems. Also presented was some more recent work that addresses the information fusion problem in a military application. From these examples, it should be clear that the success of these and related efforts is contingent upon finding effective solutions to several recurring problems. The knowledge base and diagnostic processes employed by physicians, or by tacticians, in the case of situation assessment systems, must be modeled, encoded in a computer, and validated in an objective manner. The knowledge engineering process, the encoding scheme, and the performance assessment problem all deserve considerable research attention. The design of the human-system interface is fraught with new problems, such as the requirement to explain the system's reasoning, i.e., how it reached a conclusion or recommendation. There is a pressing need for a new kind of task analysis, one that characterizes the cognitive tasking of humans in a way that identifies what should be automated. It is clear that some of the issues can only be addressed adequately within the context of an attempt to build a real system. Laboratory systems must eventually be tested in the crucible of reality. However, this trial and error approach to knowledge-based system development and evaluation is patently unsatisfactory for military applications.

The design of MI technology for military systems should probably begin with a task analysis that subserves the requirements of the knowledge engineering process; i.e., a model of the human knowledges and performances to be augmented or replaced. The adequacy of the model will depend upon its ability to deal with the issues of uncertainty in information sources and decision strategies, and the capacity requirements imposed on the humans. If a contemplated application requires humans to assess more than a few information sources, or consider more than a few rules per decision, or make more than a few decisions every minute or so, and if uncertainty in the information sources or decision rules is not significant, then automation is probably indicated. How "few" or how "uncertain?" There are no ready, generic answers to such questions. Systematic research and development that concentrates on a well-defined system application and that simultaneously attends to the issues raised here would represent an important and necessary step in the evolution of MI technology.

REFERENCES

01. Barr, A., and Feigenbaum, E. The handbook of artificial intelligence: Volume 2. Los Altos, CA: Kaufman, 1982.

02. Broadbent, D. Decision and Stress. New York, NY: Academic Press, 1971.

03. Clocksin, W., and Mellish, C. Programming in PROLOG. New York: Springer-Verlag, 1981.

04. Denning, P., Feigenbaum, E., Gilmore, P., Hearn, A., Ritchie, R., and Traub, J. The Snowbird Report: A discipline in crisis. Communications of the ACM. 1981, 370-374.

05. Dillard, R. The Demster-Shafer theory applied to tactical data fusion in an inference system. Paper presented at the Fifth MIT/ONR Workshop on Distributed Information and Decision Systems Motivated by Naval Command-Control-Communications (C3) Problems, 1982.

06. Feigenbaum, E., and McCorduck, P. The fifth generation: Artificial intelligence and Japan's computer challenge to the world. Reading, MA: Addison Wesley, 1983.

07. Godel, K. Uber formal unentscheidbare satze der Principia Mathematica und verwandter systems. Monatshefte der Mathematik und Physik, 1931, 38, 173-198.

08. Grandjean, R., Hunting, W., and Eidermann, M. VDT workstation design: Preferred settings and their effects. Human Factors, 1983, 2, 161-175.

09. Harris, S. D. Voice-controlled avionics: Programs, progress, and prognosis. Proceedings of Workshop on Standardization for Speech I/O Technology, National Bureau of Standards, Gaithersburg, MD, March 18-19, 1982.

10. Hofstadter, D. Godel, Escher, Bach: An eternal golden braid. New York, NY: Basic Books, Inc., 1979.

11. Lovelace, A. Notes upon L.F. Menabrea's sketch of the analytical engine invented by Charles Babbage. In P. Morrison and E. Morrison (Eds.) Charles Babbage and his calculating engines. New York, NY: Dover, 1961.

12. Madni, A., and Freedy, A. Decision aids for airborne intercept operations in advanced aircrafts. In Proceedings of the 17th Annual Conference on Manual Control. Los Angeles, CA: University of California, June, 1981.

13. McCarthy, J. Recursive functions of symbolic expressions and their computation by machine. Communications of the ACM, 1960, 184-195.

14. Michalski, R., Carbonnel, J. and Mitchell, T. Machine learning and artificial intelligence approach. Palo Alto, CA: Tioga Publishing Co., 1983.

15. Miller, G. The magical number seven plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97.

16. Nashelsky, L. Introduction to digital computer technology. New York, NY: Wiley, 1972.

17. Newell, A., and Simon, H. Human problem solving. Englewood Cliffs, NJ: Prentice-Hall, 1972.

18. Post, E. Formal reductions of the general combinatorial decision problem. American Journal of Mathematics, 1943, 65, 197-268.

19. Rich, E. Artificial Intelligence. New York, NY: McGraw-Hill Book Company, 1983.

20. Shortliffe, E. Computer-based medical consultations: MYCIN. New York, Elsevier, 1976.

21. Waterman, D., and Hayes-Roth, F. An investigation of tools for building expert systems. Report prepared for the National Science Foundation. Santa Monica, CA: Rand Corporation, 1982.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Special Report 84-2	2. GOVT ACCESSION NO.: A150 684
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Some critical factors that limit the effectiveness of machine intelligence technology in military systems applications	5. TYPE OF REPORT & PERIOD COVERED: Interim
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): S. D. Harris* and J. M. Owens	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Aerospace Medical Research Laboratory, Naval Air Station, Pensacola, Florida 32508-5700	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: Naval Air Systems Command 61153N WR0421001.6142
11. CONTROLLING OFFICE NAME AND ADDRESS: Naval Medical Research and Development Command, Naval Medical Command, National Capitol Region, Bethesda, Maryland 20814	12. REPORT DATE: September 1984
	13. NUMBER OF PAGES: 17
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Naval Air Systems Command, Code 330J, Washington, D. C. 20361	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

*Current mailing address for LCDR Harris is RFD 1, Box 243, Piner, Virginia 24149

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Computer Systems Design
Performance Criteria
Machine Intelligence (MI)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Recent advances in machine intelligence (MI) technology have been widely touted as a panacea for many difficult military systems problems. However, a number of factors that limit the usefulness of the technology in military applications become apparent upon perusal of the available data that reflect

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract (continued)

the performance of operational MI systems. This report serves to delineate some important problems and issues that require the attention of the research community if machine intelligence technology is to become a viable component in defense systems.

Machine intelligence technology is not the product of a recent breakthrough in computer system design as many have been led to believe. The fundamental concepts, and many of the underlying problems in the design of today's intelligent machines, were recognized fifty years ago, and earlier. Research efforts in diverse areas, some spanning decades, have culminated in the current state of the technology. However, numerous and often perplexing methodological problems currently challenge system development and thwart the widespread and successful application of the technology in critical areas.

The issues discussed in this paper arise from the failure of MI technology to confront and effectively solve many of the complex problems inherent in military systems applications. Most critical is the fact that performance criteria for intelligent machines are vague and inadequate, or completely unspecified in some instances. It is often impossible to ascertain from the reported data exactly what a system is supposed to do and how well its functions are performed. The problem of performance assessment is one of several difficult issues discussed in this report.

There are no ready solutions to the problems outlined in this report. Systematic research and development that concentrates on well-defined military systems applications and that simultaneously attends to the issues raised here would represent an important and necessary step in the evolution of MI technology.